Modeling Factors Affecting the Number of Tuberculosis Cases in Indonesia Using Geographically Weighted Generalized Poisson Regression

By

Sri Pingit Wulandari
Departement of Business Statistics, Sepuluh Nopember Institute of Technology, Surabaya, 60111, Indonesia

Mumtazah Nurul ‘Aini
Departement of Business Statistics, Sepuluh Nopember Institute of Technology, Surabaya, 60111, Indonesia

Abstract

Tuberculosis (TB) is currently still a public health problem both in Indonesia and internationally. Indonesia is ranked 2nd with the highest TB sufferers in the world. In 2020, the number of tuberculosis cases found was 351,936 cases, a decrease when compared to 2019 which was 568,987 cases. Based on the Strategic Plan of the Ministry of Health 2020, the targeted tuberculosis success rate is 90%, then nationally the success rate of tuberculosis treatment has not been achieved (82.7%). This study aims to determine the modeling of the number of tuberculosis cases and what factors have a significant effect on tuberculosis in Indonesia with the Geographically Weighted Generalized Poisson Regression (GWGPR). The results of the analysis using GWGPR shows that 12 provincial groups were formed according to variables that had a significant effect on the number of tuberculosis cases in Indonesia. There are 3 variables that have a significant effect on all provinces in Indonesia, the percentage of poor people (X1), population density (X2), and the percentage of people of productive age (X8). A variable that has no significant effect on all provinces in Indonesia is the percentage of households that have access to proper drinking water source services (X4).

Keywords: factors, Indonesia, GWGPR, public health, tuberculosis

Introduction

Indonesia is currently experiencing a double burden, on the one hand Non-Communicable Diseases (NCDs) are rising significantly, but are still faced with infectious diseases that have not yet been completed. One of the infectious diseases of concern is Tuberculosis (TB). Tuberculosis (TB) is currently still a public health problem both in Indonesia and internationally and also an important contributor to morbidity and mortality [1]. Despite being preventable and treatable, tuberculosis is a leading cause of death from a single infectious agent [2]. Tuberculosis is an infectious disease infection caused by Mycobacterium tuberculosis, a rod-shaped and acid-resistant bacterium, so it is called Acid Resistant Bacteria (BTA) and is a bacillus bacterium. which the invasion of lung tissue is the most common. Tuberculosis is difficult to treat and it is easy to produce multi-drug resistance, and it also seriously damages human tissues and organs. Consequently, tuberculosis has been one of the major infectious diseases threatening global human health [3], [4].

Indonesia is ranked 2nd with the highest TB sufferers in the world after India. Globally, an estimated 10 million people suffered from tuberculosis in 2019. Although there is a decrease in new TB cases, it is not fast enough to achieve the 2020 END TB Strategy target, namely a 20% reduction in TB cases between 2015 – 2020. In 2015 – 2019 the cumulative decrease in TB cases was only 9% [5]. In 2020 the number of tuberculosis cases found was 351,936 cases, a decrease when compared to all tuberculosis cases found in 2019, which was 568,987 cases. However, if referring to the target set by the Strategic Plan of the Ministry of Health in 2020, the tuberculosis success rate is 90%, then nationally the success rate of tuberculosis treatment has not been achieved (82.7%). This figure is lower than the previous year which reached 82.9%. The Ministry of Health noted that the success rate of treatment of TB patients has decreased since 2016. One of the strategic objectives of the Ministry of Health for 2020-2024 is to improve disease prevention and control and management of public health emergencies where one of the indicators is the decrease in TB incidence to 190 per 100,000 population by 2024. This shows that the government needs to improve health services for the treatment of TBC.

Count data can be applied to different fields, including medicine, agriculture, life science, business, social, behavioral science, and demographic and health survey data [6]. The number of tuberculosis cases in Indonesia is a data count that follows the Poisson distribution. Poisson regression is very suitable for analyzing count data if the mean and variance are the same (equidispersion). However, equidispersion conditions are difficult to meet, in general there are often cases of overdispersion where the variance is greater than the average. Generalized Poisson Regression (GPR) is one of the alternatives to overcome over/under dispersion cases in data modeling using Poisson regression. GPR is a useful model for fitting both over-dispersed and under-dispersed count data because it allows for more variability and it is more flexible in analyzing independent variables [6]. There are differences in influential factors in each topography indicating the influence of local conditions of a particular region in determining the factors that have a significant effect on tuberculosis. Therefore, this study pays attention to spatial factors so that the results of modeling the number of tuberculosis cases can describe a better relationship pattern than global regression analysis. This study will model and analyze what factors affect tuberculosis in Indonesia using Geogprahically Weighted Generalized Poisson Regression (GWGPR). This study is expected to provide an overview of the factors that affect the number of tuberculosis cases in Indonesia and can be used as input on policies for each province that will be made to reduce the number of tuberculosis diseases in Indonesia.

Previous Researches

Previous research about tuberculosis says that along with well-established risk factors (such as human immunodeficiency virus (HIV), malnutrition, and young age), emerging variables such as diabetes, indoor air pollution, alcohol, use of immunosuppressive drugs, and tobacco smoke play a significant role at both the individual and population level. Socioeconomic and behavioral factors are also shown to increase the susceptibility to infection. Specific groups such as health care workers and indigenous population are also at an increased risk of TB infection and disease [7]. Research about factors that influence current tuberculosis epidemiology says that TB has been classically associated with poverty, overcrowding and malnutrition. Low income countries and deprived areas, within big cities in developed countries, present the highest TB incidences and TB mortality rates and we must also be aware about the impact that smoking and diabetes pandemics may be having on the incidence of TB [8]. Another study by shows that global factors that influence tuberculosis sufferers are the number of HIV/AIDS sufferers, the percentage of households applying PHBS, the ratio of health education, the percentage of the population receives tuberculosis information, the number of medical personnel, the number of the population not completed elementary and the number of population graduated from high school while the influenced locally variables on the spread of tuberculosis was the percentage of healthy homes [9].

Previous research using GWPGR has been conducted by Adryanta & Purhadi (2020) concluded that the dominant variables are significant for each district/city, namely the variable percentage of households with PHBS, the percentage of healthy houses, the percentage of residents with access to proper sanitation facilities (healthy latrines), the percentage of villages/kelurahan that carry out community-based total sanitation, residents with sustainable access to drinking water quality (decent), and the percentage of poor people, which is divided into five groups [10]. Meanwhile, research by Mutfi & Ratnasari (2019) obtained the conclusion that GWGPR modeling resulted in the percentage of households with PHBS and the percentage of handling obstetric complications had a significant effect on the number of maternal deaths in all districts/cities in East Java, while the percentage of households receiving cash assistance did not have a significant effect on the number of maternal deaths in regencies/cities in East Java [11].

Thematic Maps

Thematic mapping provides today's analysts with an essential geospatial science tool for conveying spatial information [12]. Thematic maps are used to visualize and understand the spatial spread of the disease and to inform mankind. the processing of thematic data is a process with a lot of influencing challenges. The number of classes, the class definition, the classification method and the color scheme can affect the displayed map content strongly [13]. In forming a thematic map, each region has a different coloring derived from the characteristics of an area. One of the staining methods in thematic maps is the Natural Breaks Method. The Natural Breaks method is a method of grouping data patterns, where values in a class have a certain limit based on the largest range [14].

Multicholinearity

One of the conditions that must be met in the formation of a regression model with several predictor variables is that there is no case of multicholinearity or there is no correlation between one predictor variable and another predictor variable. Detection of multicollinearity cases can be carried out using the VIF Variance Inflation Factor value criterion) [15]. Multicholinearity can be suspected from the high value of the VIF or in general > 10. The VIF is expressed by Equation (1):

$$VIF = \frac{1}{1-{R_j}^2}; j = 1, 2, k \tag{1}$$

With ${R_j}^2$ is the value of the coefficient of determination between the variable *xj* and other *variables x* with Equation (2).

$${R_j}^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

Poisson Regression

Poisson regression is a type of regression analysis used to analyze discrete (count data) type response variables and sum data in which the response variable (y) follows the Poisson distribution [16][17]. Count data is a type of statistical data in which the observations can take only non-negative integer values which arise from counting rather than ranking [18]. The Poisson distribution expresses the number of events that occur in an interval of time or a certain area that does not depend on the number of events that occur in a certain time interval and region [19]. The parameters in the Poisson regression model are usually estimated using the maximum likelihood estimator (MLE). MLE suffers a breakdown when there is either multicollinearity or outliers in the Poisson regression model [20]. The probability of a Poisson Distribution y can be known by the formulation in Equation (3).

$$f(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!}\ , y = 0,1,2, \text{dan } \mu > 0 \tag{3}$$

With μ is the average and variance of the Poisson distribution. The Poisson distribution is used to model relatively rare occurrences over a selected period of time. Suppose there is a set of data with the following structure,

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & x_{i1} & \dots & x_{ip} \end{bmatrix}$$

with y_i = the i-th observation value of the response variable (Y)
x_{ip}= the *i-th* observation value of the predictor variable (X_p)
Poisson's regression model is written in Equation (4).

$$\mu_i = exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \tag{4}$$

$$\text{with } \boldsymbol{x_i} = [1 \quad x_{1i} \quad x_{2i} \quad \cdots \quad x_{ip}]^{\boldsymbol{T}}$$

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_k]^{\boldsymbol{T}}$$

Overdispersion

Overdispersed count data arise commonly in disease mapping and infectious disease studies [21]. If an overdispersion occurs in discrete data and continues to use Poisson regression as a solution method, an invalid conclusion will be obtained because the standard error value is under estimate. Overdispersion is the dispersion value of pearson chi-square and deviance divided by its degree of freedom, obtained a value greater than 0 by θ being a dispersion parameter [22].

Generalized Poisson Regression (GPR)

In reality, this assumption may not hold, either with a variance larger than the mean (overdispersion) or otherwise (underdispersion) [18, 19]. Such a violation can result in errors in decision making in hypothesis testing due to the occurrence of underestimates [23], [24]. GPR model is an appropriate model for data count in case of over/under dispersion. So that in addition μ to parameters, in GPR models there are also θ as dispersion parameters. The GPR model assumes that its random components are distributed Generalized Poisson (GP). The GP distribution is found in Equation (5) [22].

$$f(y;\mu;\theta)=\left(\frac{\mu}{1+\theta\mu}\right)^{y}\frac{(1+\theta y)^{y-1}}{y!}exp\left(\frac{-\mu(1+\theta y)}{1+\theta\mu}\right),$$

$$y=0,1,2,..$$

(5)

The average and variance of GPR models are as follows.

$E(y)=\mu$ dan $Var(y)=\mu(1+\theta\mu)^2$ (6)

If it is θ equal to 0 then the GPR model will be a regular Poisson regression model. Meanwhile, if the GPR model represents an $\theta>0$ overdispersion and vice versa if the GPR model represents an $\theta<0$ then underdispersion. The GPR model has the same shape as the Poisson regression model in Equation (7).

$\mu_i=exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\,,i=1,2,\dots,n$ (7)

The estimation of GPR parameters in Equation (7) is carried out using the MLE method with the following GPR model likelihood function.

$$L(\mu,\theta)=\prod_{i=1}^{n}f(\mu,\theta)$$

$$L(\mu,\theta)=\prod_{i=1}^{n}\left\{\left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i}\frac{(1+\theta y_i)^{(y_i-1)}}{y_i!}exp\left(\frac{-\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right)\right\}$$

$$L(\mu,\theta)=\left(\prod_{i=1}^{n}\left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i}\right)\left(\prod_{i=1}^{n}\frac{(1+\theta y_i)^{(y_i-1)}}{y_i!}\right)exp\left(\sum_{i=1}^{n}\frac{-\mu_i(1+\theta y_i}{1+\theta\mu_i}\right.$$

(8)

Then Equation (8) is changed in the form of a natural logarithm function and substitutes the value $\mu_i=exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$. Furthermore the natural logarithm equation of the likelihood function is derived against $\boldsymbol{\beta}^T$ and equated to zero to obtain the parameter $\widehat{\boldsymbol{\beta}}$. Then, to get the parameter estimator $\hat{\theta}$ then the equation is derived against and equated to zero and obtained the result θ in Equation (9).

$$\frac{ln\,L(\boldsymbol{\beta},\theta)}{\theta}=\sum_{i=1}^{n}[y_i\,exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\,(1+\theta\,exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))^{-1}+y_i(y_i-1)(1+\theta y_i)^{-1}+\Delta]$$

(9)

where $\Delta=-\,exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\,\{y_i(1+\theta\,exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))^{-1}+\Delta^*\}$

and $\Delta^*=-(1+\theta y_i)\,exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\left(1+\theta(\boldsymbol{x}_i^T\boldsymbol{\beta})\right)^{-2}$

The decrease in the ln-likelihood function against $\boldsymbol{\beta}^T$ and θ often results in an implicit equation, so that a numerical method is used, namely the Newton-Raphson iteration until a convergent parameter estimator is obtained.

GPR parameter testing is the same as testing poisson regression parameters. Simultaneous testing of GPR parameters is carried out using the MLRT method with the following hypothesis.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : at\ least\ one\ \beta_j \neq 0; j = 1,2,\cdots,k$$

Test statistics: $D(\hat{\beta}) = -2\ ln\ \Lambda = -2\ ln\left(\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right)$ (10)

Reject H₀ if the value is $D(\hat{\beta}) > \chi^2_{(\alpha,k)}$ which means that there is at least one parameter that has a significant effect on the model. Then a partial parameter test is carried out to see the significance of the parameter to the model with the following hypothesis.

$H_0 :\ \beta_j = 0$ (insignificant influence of the j-th variable)
$H_1 :$ at least one $\beta_j \neq 0$; j=1,2,…,k (the influence of the j-th variable is significant)
The test statistics used follow the z distribution i.e.

$$z = \frac{\hat{\beta}_j}{se\ (\hat{\beta}_j)} \quad (11)$$

Critical area: Reject H₀ if the value of the |Z-statistics| greater than the value $z\alpha/_2$ where α is the level of significance used.

Testing Aspect of Spatial Data

Spatial regression is one of the methods used to determine the relationship between response variables and predictor variables by paying attention to location or spatial aspects. The spatial aspect in question is that the data used has correlated errors and has spatial heterogeneity [25]. Unlike ordinary regression, the spatial regression approach considers the spatial autocorrelation among the observation. Moreover, the spatial regression models can effectively estimate the influence of independent factors on target variables by differentiating the spatial dependence by including the lag and error components of independent features [26].

Spatial data aspect testing consists of spatial dependency testing and spatial heterogeneity testing. Spatial dependency testing is performed to see if observations at one location have an effect on observations in other locations that are close together. Spatial dependency testing was performed using Moran's I test statistics with the following hypothesis [25].

$H_0 :\ I = 0$ (no spatial dependencies)
$H_1 :\ I \neq 0$ (there are spatial dependencies)
Test statistics :

$Z_{hit} = \frac{\hat{I} - E(\hat{I})}{\sqrt{Var(\hat{I})}}$	(12)

where,
$\hat{I}$: moran index i.e. $\hat{I} = \frac{e^TWe}{e^Te}$
Z_{hit} : statistical value of Moran index test

W : spatial weighting matrix
$E(\hat{I})$: the expectation value of the Moran index is $E(\hat{I}) = \frac{tr(MW)}{(n-k)}$
$Var(\hat{I})$: the standard deviation of the Moran index is

$$Var(\hat{I}) = \frac{\left[tr(\boldsymbol{MWMW^T}) + tr(\boldsymbol{MW})^2 + \left(tr(\boldsymbol{MW})\right)^2\right]}{d - E(\hat{I})^2} \qquad (13)$$

$$\text{with, } d = (n-k)(n-k+2)$$

$$M = (I - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T).$$

Decision: Reject H$_0$ if the value |Z-statistics| $> Z_{\alpha/2}$ or *p-value* < α, which means there are spatial dependencies in the model.

Spatial heterogeneity testing is carried out to see if there are any peculiarities at each observation site, so that the resulting regression parameters vary spatially. Spatial heterogeneity testing was carried out using Breusch-Pagan test statistics with the following hypothesis.

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (variance between locations is the same)
H_1 : there is at least one $\sigma_i^2 \neq \sigma^2$ (variance between locations is different)
Test statistics

$$: BP = (1/2)\boldsymbol{f^T Z}(\boldsymbol{Z^T Z})^{-1}\boldsymbol{Z^T f} \qquad (14)$$

Where

$$f = (f_1, f_2, \dots, f_n)^T \text{with } f_i = \left(\frac{\hat{e}_i^2}{\hat{\sigma}^2} - 1\right)$$

With $\hat{e}_i$ is the residual of the Ordinary Least Square (OLS) method for the i-th observation, **Z** is a $nx(k+1)$ sized matrix containing vectors that are already in the standard normal for each observation.

Decision: Reject H$_0$ if the value of the $BP > \chi^2_{(\alpha,k-1)}$ or *p-value* < α which means that heteroskedasticity occurs in the model.

Geographically Weighted Generalized Poisson Regression (GWGPR)

The Geographically Weighted Generalized Poisson Regression (GWGPR) model is a GPR regression development model that uses geographical weighting, specifically latitude and longitude points, in parameter estimation, resulting in different parameter estimates per region. The probability distribution function of the GWGPR for each location is contained in Equation (15).

$$f(y_i|\mu_i,\theta) = \left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} exp\left(\frac{-\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right) \qquad (15)$$

Where $\boldsymbol{\mu_i}$

$$= exp\ \left(\sum_{j=0}^{k} \beta_j(u_i, v_i)\ x_{ji}\right), \text{(16)}$$

$$\text{so } \boldsymbol{y_i} \sim exp\ \left(\sum_{j=0}^{k} \beta_j(u_i, v_i)\ x_{ji}\right) \text{(17)}$$

The form of the GWGPR equation is found in Equation (18).

$$\boldsymbol{\mu_i} = e^{X_i^T \beta(u_i, v_i)}$$
$$\boldsymbol{\mu_i} = exp\ (\beta_0(u_i, v_i) + (\beta_0(u_i, v_i)x_{i1} + \cdots + (\beta_k(u_i, v_i)x_{ik}$$
(18)

Where k is the multiplicity of predictor variables, y_i is the observation value of the i-th response, x_{ji} is the observed value of the j-th predictor variable on the observation of location (u_i, v_i), $\beta_j(u_i, v_i)$ is the regression coefficient of the j-th predictor variable for each location (u_i, v_i), and (u_i, v_i) is the latitude and longitude coordinates of the i-th point at a geographical location.

GWGPR models require graphics weighting and bandwidth in their parameter estimation. Theoretically, bandwidth is a circle with a radius of b from the center point of the location which is used as the basis for determining the weight of each observation of the regression model at that location. The selection of the optimum bandwidth is very important because it will affect the accuracy of the model to the data, namely regulating the variance and bias of the model. The determination of the optimum method was carried out using the Cross Validation (CV) method with Equation (19) [27].

$$CV(b) = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(b)\right)^2$$
(19)

By $\hat{y}_{\neq i}(b)$ being an estimator y_i value with location (u_i, v_i) observation is omitted from the estimation process.

In the process of estimating the parameters of the GWGPR model at a point it is necessary to have a spatial weighting where the weighting used is a kernel function. (u_i, v_i). In this study, the Adaptive Bisquare kernel function was used with the formula in Equation (20).

$$w_{ij}(u_i, v_i) = \begin{cases} \left(1 - \left({d_{ij}}/{b_{i(p)}}\right)^2\right)^2 & \text{, untuk } d_{ij} \leq b_{i(p)} \\ 0 & \text{, untuk } d_{ij} > b_{i(p)} \end{cases}$$
(20)

with is $d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$ the Euclidean distance between location (u_i, v_i) and location $\left(u_j, v_j\right)$, and b is the optimum bandwidth value at each location.

The estimation of the parameters of the GWGPR model is carried out using the MLE method, namely by maximizing the likelihood function [27]. The first step of this method is to form the likelihood function. The likelihood function of the GWGPR model is found in Equation (21).

$$L(\theta, \boldsymbol{\beta}(u_i, v_i), i = 1,2,\dots,n) = \prod_{i=1}^{n} f(y_i)$$
$$L(\theta, \boldsymbol{\beta}(u_i, v_i), i = 1,2,\dots,n) = \prod_{i=1}^{n} \left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i} \frac{(1+\theta y_i)^{y_i-1}}{y_i!} exp\left(\frac{-\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right) \tag{21}$$

In the GWGPR model, the factors that are considered as weights are the geographical factors of each observation location. The natural logarithm function (ln) likelihood based on geographical factors is found in Equation (22).

$$\begin{aligned} &\ln\left(\left(L^*\left((\theta, \boldsymbol{\beta}(u_i, v_i)\right)\right)\right) \\ &= \sum_{i^*=1}^{n} w_{ii^*}\left[\, y_{i^*}\boldsymbol{x}_{i^*}^T\boldsymbol{\beta}(u_i, v_i)\right] + \sum_{i^*=1}^{n} w_{ii^*}\left[- y_{i^*}\, ln\left(1+\theta e^{\boldsymbol{x}_{i^*}^T\boldsymbol{\beta}(u_i,v_i)}\right)\right] \\ &+ \sum_{i^*=1}^{n} w_{ii^*}\left[(y_{i^*}-1)\ln(1+\theta\, y_{i^*}) - \ln(\, y_{i^*}!)\right] \\ &+ \sum_{i^*=1}^{n} w_{ii^*}\left[\frac{e^{\boldsymbol{x}_{i^*}^T\boldsymbol{\beta}(u_i,v_i)}(1+\theta\, y_{i^*})}{1+\theta e^{\boldsymbol{x}_{i^*}^T\boldsymbol{\beta}(u_i,v_i)}}\right] \end{aligned} \tag{22}$$

Furthermore, a Newton Rhapson iteration was carried out in order to obtain a parameter estimator of the GWGPR model that closes the form until it iterates to m = m + 1 the $\widehat{\boldsymbol{\beta}}^*_{(\mathrm{m})}$ convergent value, that is, $\left\|\widehat{\boldsymbol{\beta}}^*_{(\mathrm{m}+1)} - \widehat{\boldsymbol{\beta}}^*_{(\mathrm{m})}\right\| < \varepsilon$ whereε it is a very small number of 10^{-6}.

One method of simultaneously testing model parameters can use the Maximum Likelihood Ratio Test (MLRT) to test the significance of geographical or location factors. The hypotheses used are as follows [28].

Simultaneous Testing of GWGPR models

Simultaneous testing was carried out using mlrt with the following hypothesis.

$$H_0 : \beta_1(u_1, v_1) = \beta_2(u_2, v_2) = \cdots = \beta_k(u_i, v_i) = 0; i = 1,2,\dots,n$$
$$H_1 : \text{there is at least one } \beta_k(u_i, v_i) \neq 0; i = 1,2,\dots,n$$

The test statistics used are *likelihood* ratio statistics in Equation (23).

$$D\left(\hat{\beta}(u_i, v_i), i = 1,2,\dots,n\right) = -2\ln(\Lambda) = -2ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \tag{23}$$

Decision: Reject H_0 if the value $D\left(\hat{\beta}(u_i, v_i)\right) > \chi^2{}_{(\alpha,db)}$ which means that there is at least one predictor variable in the GWGPR model that has a significant effect on the response variable.

Partial Parameter Testing

Partial parameter testing is carried out to find out which parameters have a significant effect on the response variables at each location with the following hypothesis.

H_0 : $\beta_j(u_i, v_i) = 0$ (variable x_{ji} has no significant effect)
H_1 : minimal ada satu $\beta_j(u_i, v_i) \neq 0$; j=1.2. k; i=1.2,...,n
The test statistics used are z test statistics, in Equation (24).

$$z = \frac{\hat{\beta}_j(u_i, v_i)}{se\left(\hat{\beta}_j(u_i, v_i)\right)} = \frac{\hat{\beta}_j(u_i, v_i)}{\sqrt{Var\left(\hat{\beta}_j(u_i, v_i)\right)}} \quad (24)$$

The area of rejection is H₀ will be rejected if the value of |Z-statistics| greater than $z\alpha/_2$, where α is the level of significance used.

Tuberculosis

Tuberculosis is an infectious disease infection caused by Mycobacterium tuberculosis, a rod-shaped and acid-resistant bacterium, so it is called Acid Resistant Bacteria (BTA) and is a bacillus bacterium that is so strong that it takes a long time to treat it. Tuberculosis can affect anyone, especially the population of productive age or who are still actively working, namely the age of 15-50 years. In almost all cases, tuberculosis infection is acquired through the inhalation of fairly small particles of germs (about 1-5 μm). Droplets are secreted during coughing, laughing, or sneezing. A pulmonary-infected nucleus can occur, the inhaled organism must first fight the pulmonary defense mechanisms and enter the lung tissue [29].

People from low socioeconomic-status populations are known to be at high risk of becoming sick from tuberculosis, and in low-burden tuberculosis countries, substantial declines in tuberculosis morbidity and mortality have occurred as a result of improvement in overall living conditions [30]. indoor air pollution, living in a house with a low number of windows per room, and socioeconomic position of the household, can be powerful predictors of tuberculosis infection and disease [31]. Furthermore, people who have had one episode of tuberculosis are at increased risk of developing tuberculosis again, further exacerbating the vicious cycle of poverty and tuberculosis. Addressing socioeconomic factors, including smoking and indoor air pollution, could be just as important as addressing host and pathogen factors in easing the global burden of tuberculosis. A controlled human infection model to improve the understanding tuberculosis infection is an unmet need in the field [32].

Method

Data Source

The data used in this study are secondary data obtained from the Indonesian Statistical Publication obtained from the website of the Central Statistics Agency of the Republic of Indonesia (www.bps.go.id), the publication of the Indonesian Health Profile 2021 obtained from the website of the Ministry of Health of the Republic of Indonesia (www.kemkes.go.id), and data on the coordinates of Indonesia's longitude latitude obtained from the Github website (https://github.com). This research unit is 34 provinces in Indonesia.

Research Variable

The variables used in this study are research variables from object data, namely 34 Indonesian provinces in Table 1.

Table 1. *Research Variable*

Variable	Information	Unit
Y	Number of cases of tuberculosis disease	Case
X_1	Percentage of poor population	Percent
X_2	Population density	Soul/km^2
X_3	Percentage of health complaints	Percent
X_4	Percentage of households that have access to adequate drinking water source services	Percent
X_5	Percentage of households that have access to proper sanitation	Percent
X_6	Percentage of livable houses	Percent
X_7	Percentage of malnutrition of the community	Percent
X_8	Percentage of the population of productive age	Percent
X_9	Percentage of health workers	Percent
X_{10}	Percentage of places and public facilities that are supervised according to standards	Percent
X_{11}	Percentage of food management places that meet the requirements according to standards	Percent
X_{12}	Percentage of residents who smoke	Percent
u	Latitude coordinates	-
v	Longitude coordinates	-

Analysis Method

This study will analyze what variables are suspected to affect the number of tuberculosis cases in Indonesia using the Geographically Weighted Generalized Poisson Regression (GWGPR) method because the number of tuberculosis cases in Indonesia is a data count that follows the Poisson distribution. Poisson regression is very suitable for analyzing count data. The Geographically Weighted Generalized Poisson Regression (GWGPR) model is a GPR regression development model but uses geographical weighting, namely latitude and longitude points in the parameter estimation, so that different parameter estimates will be generated for each region. The analysis steps carried out in this study are GWGPR as follows.

a. Describe the characteristics of tuberculosis disease data along with the factors that affect tuberculosis disease in provinces in Indonesia.
b. Create a thematic map of provinces in Indonesia using the Natural Breaks method.
c. Perform a multicholinearity analysis on predictor variables by looking at vif values.
d. Perform a Poisson regression analysis with the following steps.
e. Perform parameter estimation of the Poisson regression model
f. Perform signification testing of Poisson regression model parameters simultaneously
g. Perform partial poisson regression model parameter signification testing
h. Perform the dispersion test of the Poisson regression model
i. Perform a GPR analysis with the following steps.
j. Perform parameter estimation of GPR models
a. Perform simultaneous signification testing of GPR model parameters
k. Perform partial GPR model parameter signification testing
l. Test aspects of spatial data to see spatial heterogeneity of data
m. Perform a GWGPR analysis with the following steps.
n. Calculating euclidean distances between observation sites based on geographical position.
o. Determine the optimum bandwidth by using Cross Validation (CV)

p. Calculate the weighting matrix by using the Adaptive Bisquare Kernel weighting function
q. Perform parameter estimation of the GWGPR model
r. Perform simultaneous and partial GWGPR model parameter significance tests on regions where model parameters have been estimated
s. Perform the formation of a grouping map
t. Interpreting the results obtained
u. Draw conclusions and suggestions.

Result and discussion

Data Characteristics and Map of Distribution of Research Variables

Data characteristics of the number of tuberculosis cases in each province in Indonesia and 12 predictor variables that are suspected to affect the number of tuberculosis cases in Indonesia are contained in Table 2.

Table 2. Data Characteristics

Variable	Median	Variance	Min	Max
Y	4490	250,161,857	918	79,423
X_1	8.735	29.569	3.780	26.64
X_2	103	7,338,488	9	15,907
X_3	28.02	38.81	15.97	44.00
X_4	87,95	91.96	62.47	99.84
X_5	79.90	99.12	40.31	96.96
X_6	58,83	159.52	28.56	86.19
X_7	1.150	0.420	0.100	2.900
X_8	68.964	2.024	64.440	71.566
X_9	0.01901	0.00094	0.00451	0.11734
X_{10}	55.55	768.72	0.00	94.60
X_{11}	45.45	169.80	12.70	70.00
X_{12}	27.74	8.236	20.500	33.430

Table 2 shows that the middle value of the number of tuberculosis cases in Indonesia in 2020 has a middle value of 4,490 cases with a large variance of 250,161,657. The magnitude of the variance value indicates that there are many diversity or differences in tuberculosis cases in each province in Indonesia. The highest number of tuberculosis cases in 2020 was 79,423 cases, namely in West Java Province, while the lowest number of tuberculosis cases in 2020 was 918 cases, namely in North Kalimantan Province.

The distribution of data based on each research variable is displayed in the form of a thematic map with the aim that the distribution of data from each variable can be known easily. The map of the distribution of the number of tuberculosis cases in Indonesia is found in Figure 1.

Figure 1 shows that the distribution of the number of tuberculosis cases in Indonesia is divided into three categories, namely low (918-9,600 cases), medium (9,601-24,274 cases), and high (24,275-79,423 cases). The provinces that have a relatively high number of tuberculosis cases (24,275-79,423 cases) are West Java Province, East Java Province, and Central Java Province. Meanwhile, the provinces that have a relatively low number of tuberculosis cases (918-9,600 cases) are Aceh, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Bangka Belitung Islands, Riau Islands, In Yogyakarta, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, and Papua.

Patterns of Relationships Between Variables

The pattern of the relationship between the number of cases of tuberculosis disease and the factors affecting it needs to be identified before modeling. Identification of the pattern of relationship between the response variables of the number of tuberculosis disease cases and the factors affecting it was carried out using the scatter plot in Figure 2.

Figure 2 shows the pattern of the relationship between the number of tuberculosis cases in Indonesia in 2020 with factors suspected to have an effect It is visually seen that the relationship between the number of tuberculosis cases (Y) and population density (X_2), the percentage of health complaints (X_3), the percentage of households that have access to adequate drinking water source services (X_4), the percentage of the population of productive age (X_8), the percentage of health workers (X_9), and the percentage of the population who smoke (X_{12}) tend to have a positive linear relationship which means it is directly proportional to the number of tuberculosis cases. Meanwhile, the variables of the percentage of poor people (X_1), the percentage of households that have access to proper sanitation (X_5), the percentage of households occupying livable houses (X_6), the percentage of malnutrition of the community (X_7), the percentage of Public Places and Facilities (TFU) carried out supervision according to standards (X_{10}), and the percentage of Food Management Sites (TPP) that meet the requirements according to the standard (X_{11}) tends not to have a linear relationship. This can be caused by an outlier in the data.

Multicholinearity

Detection of multicollinearity cases is carried out using the criteria of the value of VIF (*Variance Inflation Factor*). The VIF values on each predictor variable are shown in Table 3.

Table 3. *VIF Values*

Variable	VIF Values	Variable	VIF Values
X_1	2.36	X_7	2.10
X_2	9.85	X_8	4.10
X_3	2.22	X_9	3.86
X_4	6.68	X_{10}	3.01
X_5	5.17	X_{11}	2.64
X_6	5.49	X_{12}	3.34

Table 3 informs that all predictor variables have met the non-multicholinearity assumption because the VIF value of the 12 predictor variables <10. This suggests that no predictor variable correlates with each other with other predictor variables.

Overdispersion Examination Poisson Regression

Detecting the presence of overdispersion can be done by looking at the devians value in the Poisson regression model divided by the degree of freedom, with the results contained in Table

Table 4. *Overdispersion*

Devians	db	θ
38,352.1217	21	1,826.915

Table 4 shows that the devians value was obtained at 38,352.1217 with a free degree of 21. The value of the quotient of the Devians of the Poisson regression model with free degree is 1,826.915. The quotient value is greater than 0 which indicates the occurrence of a case of overdispersion or a variance value greater than the average value so to overcome it can use Generalized Poisson Regression (GPR).

Modeling the Number of Tuberculosis Cases in Indonesia in 2020 Using GPR

One method that can be used to analyze data that occurs in cases of overdispersion is Generalized Poisson Regression (GPR). Estimation and testing of GPR model parameters are contained in Table 5.

Table 5. *Estimated GPR Model Parameters*

Parameter	Estimation	|Z-statistics|	$Z_{0,05/2}$	*P-value*	Decision
β_0	39.321	1.06	1.96	0.2952	Failed to Reject H0
β_1	-0.00973	-0.49	1.96	0.6307	Failed to Reject H0
β_2	0.00717	1.62	1.96	0.1155	Failed to Reject H0
β_3	-0.01731	-1.3	1.96	0.2026	Failed to Reject H0
β_4	0.02512	2	1.96	0.0539	Failed to Reject H0
β_5	-0.01547	-1.2	1.96	0.2372	Failed to Reject H0
β_6	-0.01691	-2.12	1.96	0.0418	Reject H0
β_7	0.1112	0.63	1.96	0.5315	Failed to Reject H0
β_8	0.0446	0.84	1.96	0.4062	Failed to Reject H0
β_9	747.962	5.76	1.96	<0.0001	Reject H0
β_{10}	0.005339	1.35	1.96	0.1852	Failed to Reject H0
β_{11}	-0.01455	-2.32	1.96	0.0263	Reject H0
β_{12}	0.0301	0.87	1.96	0.393	Failed to Reject H0
Deviance	1,040,959				

Simultaneous testing of GPR model parameters aims to determine whether the predictor variable has a significant influence on the response variable simultaneously. The hypothesis on simultaneous testing is as follows.

Hypothesis:
H0: ($\beta_1 = \beta_2 = \cdots = \beta_{12} = 0$All predictor variables have no significant effect on the number of tuberculosis cases in Indonesia)
H1: there is at least one (There is at least 1 predictor variable $\beta_j \neq 0;\ \ j = 1,2,\cdots,12$ that has a significant effect on the number of tuberculosis cases in Indonesia)
Significant level: $\alpha = 0.05$
Critical area: Reject H0 if the value $D(\hat{\beta}) > X^2_{(\alpha,k)}$

The test statistics using the devians value contained in Table 5 showed that the test statistical value of $D(\hat{\beta})$ is 1,040,959 was greater than $X^2_{(0.05;12)}$ = 21.02, so a decision was made to reject H$_0$. This means that there is at least one predictor variable that has a significant effect on the number of tuberculosis cases in Indonesia.

In order to find out which variables have an effect, it can be continued on a partial test with the following hypothesis.

Hypothesis:

$H_0 : \beta_j = 0$ (The j-th variable has no significant effect on the number of tuberculosis cases in Indonesia)
$H_1 :$ *at least one* $\beta_j \neq 0$; j=1. 2.,12 (The j-th variable has a significant effect on the number of tuberculosis cases in Indonesia)
Significant level: $\alpha = 0.05$
Critical area: Reject H$_0$ if the value and $Z_{hitung} > Z_{\alpha/2}$*p-value* $< \alpha$

The statistics test uses the |Z-statistics| values found in Table 4. The parameter is said to have a significant effect on the model (reject H$_0$) if the value of |Z-statistics| > Z$_{0.05/2}$ is reinforced by *the p-value* value of < α. With a significance level of 5% obtained a Z value of $_{0.05/2}$ of 1.96 and *a p-value* of < α = 0.05, the parameters that have a significant effect on the model are the percentage of households occupying livable houses (X_6), the percentage of health workers (X_9), and the percentage of Food Management Sites (TPP) that meet all standards (X_{11}). The formed GPR model is as follows.

$\hat{\mu}$ =exp (3.9321 − 0.00973 X$_1$ + 0.00717 X$_2$ − 0.01731 X$_3$ + 0.02512 X$_4$ − 0.01547 X$_5$ − 0.01691 X$_6$ + 0,1112 X$_7$ + 0.0446 X$_8$ + 74.7962 X$_9$ + 0.005339 X$_{10}$ − 0.01455 X$_{11}$ + 0,03010 X$_{12}$)

Testing Aspects of Spatial Data

Spatial aspect testing is carried out to find out whether there are spatial aspects in the data or not. Testing aspects of spatial data can be done by conducting spatial dependency testing and spatial heterogeneity testing. Spatial dependency testing was performed using Moran's I test statistics with the following hypothesis.

Hypothesis:

$H_0 : I = 0$ (no spatial dependencies)
$H_1 : I \neq 0$ (there are spatial dependencies)
Significant level: $\alpha = 0.05$
Critical Area: Reject H$_0$ if the value is p-value<α

Based on the test results using R software, a p-value of 0.00 is smaller than the significance level of 5%, so it fails to reject H$_0$ which means there are spatial dependencies between regions. Then, spatial heterogeneity testing was carried out using Breusch-Pagan (BP) test statistics with the following hypothesis.

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_{34}^2 = \sigma^2$ (variance between locations is the same)

H_1 : there is at least one (variance between different locations)$\sigma_i^2 \neq \sigma^2, i = 1,2,\ldots,34$
Significant level: $\alpha = 0.05$
Decision: Reject H₀ if the value of > and $BP\chi^2_{(\alpha,k-1)}$*p-value*<α
Test statistics:

Table 6. *Spatial Heterogeneity Testing Results*

BP	***k***	$X^2_{(0,05;11)}$	***p-value***
21,827	12	19,67	0,02

Table 6 shows that the statistical value of the *BP* test of 21.827 is greater than the value of$X^2_{(0.05;11)}$ 19.67 reinforced by a p-value of 0.02 less than $\alpha = 0.05$ so it was decided to reject H₀ which means that the variance between locations is different or there is spatial heterogeneity in the faithfulp The observation location is a province in Indonesia and the analysis can be continued using the GWGPR method. This study uses GWGPR or spatial points because they want to find out more about the condition of each region.

Modeling The Number Of Tuberculosis Cases In Indonesia In 2020 Using Gwgpr

Data on the number of tuberculosis cases in Indonesia occurs cases of overdispersion and there are spatial influences. Thus, the appropriate analysis method used is the Geographically Weighted Generalized Poisson Regression (GWGPR) method. Modeling the number of tuberculosis cases in Indonesia with GWGPR consists of testing aspects of spatial data, estimating parameters, testing the significance of parameters simultaneously, and testing the significance of parameters partially. This study uses Adaptive Bisquare Kernel because the use of this method for determining bandwidth values will be suitable for scattered observations with irregular and grouped patterns. The Adaptive Bisquare Kernel method makes it possible to obtain different bandwidth values for all observation points. This is because the Adaptive Bisquare Kernel method can adjust to the conditions of the observation point.

Simultaneous testing of the parameters of the GWGPR model aims to determine whether the predictor variable has a significant influence on the response variable simultaneously. The hypothesis on simultaneous testing is as follows.

Hypothesis:

H₀: $\beta_1(u_i,v_i) = \beta_2(u_i,v_i) = \cdots = \beta_{12}(u_i,v_i) = 0$; i=1, 2,34

(All predictor variables have no significant effect on the number of tuberculosis cases in Indonesia)

H₁: there is at least one$\beta_j(u_i,v_i) \neq 0$; i=1, 2,34 (There is at least 1 predictor variable that has a significant effect on the number of tuberculosis cases in Indonesia)
Significant level: $\alpha = 0.05$
Critical area: Reject H₀ if the value and $D(\hat{\beta}) > X^2_{(\alpha,df)}$*p-value* < α
Test Statistics:

Table 7. *GWGPR Simultaneous Testing Results*

$D(\hat{\beta})$	**df**	$X^2_{(0,05;12)}$
2,131,543.55	12	21.02

Table 7 shows that the statistical value of the test of $D\left(\hat{\beta}\right) = 2{,}131{,}543.55$ is greater than $X^2_{(0.05;12)}$ the amount of 21.02, so a decision was made to reject H_0. This means that there is at least one parameter of the GWGPR model that has a significant effect on the number of tuberculosis cases in Indonesia. In order to find out which variables have an effect, it can be continued on a partial test with the following hypothesis.

Hypothesis:

$$H_0 : \beta_j(u_i, v_i) = 0$$

(Variable x_{ji} has no significant effect on the number of tuberculosis cases in Indonesia)

$$H_1 : \text{ minimal ada satu } \beta_j(u_i, v_i) \neq 0;j=1,2,...,12;i=1,2,...,34$$

(Variable x_{ji} has a significant effect on the number of tuberculosis cases in Indonesia)

Based on the calculation results, a different |Z-statistics| value is obtained at each location. Test statistics can be viewed based on the value of the |Z-statistics| each of the parameters of the predictor variable compared to $Z_{0.05/2}$ which is 1.96. When the value is |Z-statistics|> 1.96 then Reject H_0, which means a significant parameter to the model. The results of significant parameters were then grouped to find out the mapping of the entire province. The grouping results show that there are three variables that have a significant effect on all provinces in Indonesia, namely the percentage of poor people (X_1), population density (X_2), and the percentage of people of productive age (X_8). A variable that has no significant effect for all provinces in Indonesia is the percentage of households that have access to adequate drinking water source services (X_4). The variable percentage of community malnutrition (X_7) has a significant effect in 33 provinces in Indonesia. The variable percentage of health workers (X_9) has a significant effect in 27 provinces in Indonesia. The variable percentage of public places and facilities carried out supervision according to standards (X_{10}) has a significant effect in 22 provinces in Indonesia. The variables of the percentage of households that have access to proper sanitation (X_5), the percentage of livable houses (X_6), and the percentage of food management sites that meet the requirements according to the standard (X_{11}) have a significant effect in 24 provinces in Indonesia. The variable percentage of the population smoking (X_{12}) had a significant effect in 8 provinces in Indonesia. The variable percentage of health complaints (X_3) has a significant effect in 4 provinces in Indonesia. Mapping of regions based on significant variables in Figure 3.

Figure 3 shows that there are 12 groups with adjacent regions that tend to have similar characteristics. After obtaining significant variables from the results of partial parameter testing, an example model was presented in one of the locations, namely North Sulawesi Province. The results of estimating GWGPR parameters and significant parameters in North Sulawesi Province show that the percentage of poor people (X_1), population density (X_2), percentage of households that have access to proper sanitation (X_5), percentage of livable houses (X_6), percentage of the population of productive age (X_8), percentage of health workers (X_9), and the percentage of the population smoking (X_{12}) had a significant effect on the model. The GWGPR model formed in North Sulawesi Province is as follows.

$$\hat{\mu}=\exp(0.001993 + 0.028418\ X_1 - 0.000051\ X_2 + 0.001855\ X_3 - 0.003787\ X_4 + 0.024064\ X_5 - 0.013314\ X_6 - 0.046415\ X_7 + 0.129289\ X_8 - 0.00076\ X_9 - 0.000247\ X_{10} - 0.000569\ X_{11} + 0.019703\ X_{12})$$

The model can be interpreted, one of which isthat every increase in the percentage of the population who smokes (X_{12}) by 1%, it will increase the number of tuberculosis cases by *exp* (0.019703) = 1.02 times the previous number assuming the other variables are constant.

Conclusion

The conclusions obtained based on the results of the analysis that has been carried out are as follows.

1. The results of the data characteristics show that there are many diversity or differences in tuberculosis cases in each province in Indonesia. The highest number of tuberculosis cases in 2020 was 79,423 cases, namely in West Java Province, while the lowest was 918 cases in North Kalimantan Province.
2. Modeling using GWGPR shows that 12 provincial groups were formed according to variables that had a significant effect on the number of tuberculosis cases in Indonesia. There are three variables that have a significant effect on all provinces in Indonesia, namely the percentage of poor people (X1), population density (X2), and the percentage of people of productive age (X8). A variable that has no significant effect for all provinces in Indonesia is the percentage of households that have access to adequate drinking water source services (X4).

Credit

Conceptualization, Methodology, Writing - original draft preparation, and Supervision: Sri Pingit Wulandari, Mumtazah Nurul ’Aini; Formal analysis and investigation: Sri Pingit Wulandari, Mumtazah Nurul ’Aini; Writing - review and editing: Sri Pingit Wulandari, Mumtazah Nurul ’Aini; Funding acquisition: Sri Pingit Wulandari, Mumtazah Nurul ’Aini; Resources: Sri Pingit Wulandari, Mumtazah Nurul ’Aini

References

C. Abbafati et al., “Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019,” Lancet, vol. 396, no. 10258, pp. 1204–1222, 2020, doi: 10.1016/S0140-6736(20)30925-9.

H. H. Kyu et al., “Global, regional, and national burden of tuberculosis, 1990-2016: Results from the Global Burden of Diseases, Injuries, and Risk Factors 2016 Study,” Lancet Infect. Dis., vol. 18, no. 12, pp. 1329–1349, 2018, doi: 10.1016/S1473-3099(18)30625-X.

N. Chu, Q. Q. Gao, and C. C. Zhou, “Trends in Spatiotemporal Distribution of Pulmonary Tuberculosis Morbidity and Mortality in China: 2004–2018 (in Chinese),” Chinese J. Public Heal., vol. 38, no. 2022, pp. 887–890, 2022.

Y. H. Sun, M. Z. Tian, and Y. W. Nie, “Application of Spatial Panel Data Model in the Analysis of National Tuberculosis Surveillance Data From 2015 to 2019 (In Chinese),” Chinese Prev. Med., vol. 23, no. 2022, pp. 436–441, 2022.

WHO, “Global Tuberculosis Report 2021,” Geneva, 2021.

J. U. Ibeji, T. Zewotir, D. North, and L. Amusa, “Modelling Fertility Levels in Nigeria using Generalized Poisson Regression-Based Approach,” Sci. African, vol. 9, 2020, [Online]. Available: https://doi.org/10.1016/j.sciaf.2020.e00494

P. Narashiman, J. Wood, C. R. MacIntyre, and M. Dilip, “Risk Factors for Tuberculosis,” Pulm.

Med., vol. 2013, 2013.

J.-P. Millet et al., “Factors that Influence Current Tuberculosis Epidemiology,” Eur. Spine J., vol. 22, no. 2013, pp. 539–548, 2013.

H. Khaulasari, “Modelling Mixed Geographically Weighted Poisson Regression for Tuberculosis Disease in Surabaya,” J. Phys. Conf. Ser., vol. 1490, 2020.

M. Adryanta and P. Purhadi, “Analisis Metode Geographically Weighted Generalized Poisson Regression untuk Pemodelan Faktor yang Mempengaruhi Jumlah Kematian Anak di Provinsi Jawa Timur,” J. Sains dan Seni ITS, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.43562.

I. A. Mutfi and V. Ratnasari, “Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kematian Ibu di Jawa Timur Menggunakan Geographically Weighted Generalized Poisson Regression,” J. Sains dan Seni ITS, vol. 7, no. 2, 2019, doi: 10.12962/j23373520.v7i2.34091.

B. T. Fraser and R. G. Congalton, “Evaluating the Effectiveness of Unmanned Aerial Systems (UAS) for Collecting Thematic Map Accuracy Assessment Reference Data in New England Forests,” Forest, vol. 10, no. 24, 2019, [Online]. Available: https://doi.org/10.3390/f10010024

C. Juergens, “Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps,” J. Cartogr. Geogr. Inf., vol. 70, no. 155–161, 2020, [Online]. Available: https://doi.org/10.1007/s42489-020-00057-w

A. Basofi, A. Fariza, A. Ahsan, and I. Kamal, “A Comparison between Natural and Head/Tail Breaks in LSI (Landslide Susceptibility Index) Classification for Landslide Susceptibility Mapping : A Case Study in Ponorogo, East Java, Indonesia,” Int. Conf. Sci. Inf. Technol., 2015.

R. R. Hocking and A. Dean, Methods and Applications of Linear Models, vol. 39, no. 3. Canada: John Wiley & Sons, 2003. doi: 10.2307/1271138.

[16] A. Agresti, Categorical Data Analyst Second Edition. New Jersey: John Wiley & Sons, 2002.

D. N. Sari, P. Purhadi, S. P. Rahayu, and I. Irhamah, “Estimation and Hypothesis Testing for the Parameters of Multivariate Zero Inflated Generalized Poisson Regression Model.,” Symmetry 2021, vol. 13, 2021.

A. C. Cameron and P. K. Trivedi, Regression Analysis of Count Data, 2nd ed. USA: Cambridge University Press, 2013.

R. E. Walpole, Introduction to Statistics. Jakarta: Gramedia Pustaka Utama, 2017. doi: 10.5005/jp/books/13016_12.

K. C. Arum, F. I. Ugwuowo, and H. e. Oranye, “Robust Modified Jackknife Ridge Estimator For the Poisson Regression Model with Multicollinearity and Outliers,” Sci. African, vol. 17, 2022, [Online]. Available: https://doi.org/10.1016/j.sciaf.2022.e01386

F. Mutiso, J. L. Pearce, S. E. Benjamin-Neelon, N. T. Mueller, H. Li, and B. Neelon, “Bayesian Negative Binomial Regression with Spatially Varying Dispersion: Modeling COVID-19 Incidence in Georgia,” Spat. Statictics, vol. 52, 2022, [Online]. Available: https://doi.org/10.1016/j.spasta.2022.100703

F. Famoye, J. Wulu, and K. Singh, “On the Generalized Poisson Regression Model with an Application to Accident Data,” J. Data Sci., vol. 2, no. 3, pp. 287–295, 2021, doi: 10.6339/jds.2004.02(3).167.

L. Zha, D. Lord, and Y. Zou, “The Poisson Inverse Gaussian (PIG) Generalized Linear Regression Model for Analyzing Motor Vehicle Crash Data,” J. Transp. Saf. Secur., vol. 8, 2016.

J. M. Hilbe, Poisson Inverse Gaussian Regression. In Modeling Count Data. New York: Cambridge University Press, 2014.

L. Anselin, Spatial Econometrics : Method and Models. Netherlands: Kluwer Academic Publisher, 1988.

S. Sannigrahi, F. Pilla, B. Basu, A. S. Basu, and A. Molter, "Examining The Association between Socio-Demographic Composition and COVID-19 Fatalities in The European Region Using Spatial Regression Approach," Sustain. Cities Soc., vol. 62, 2020.

T. Nakaya, A. S. Fotheringham, C. Brunsdon, and M. Charlton, "Geographically Weighted Poisson Regression for Disease Association Mapping," Stat. Med., vol. 24, no. 17, pp. 2695–2717, 2005, doi: 10.1002/sim.2129.

R. E. Caraka and H. Yasin, Geographically Weighted Regression (GWR). Yogyakarta: MOBIUS, 2014. doi: 10.4135/9781412953962.n81.

WHO, "Global Tuberculosis Control Report," Geneva, 2011.

K. F. Ortblad, J. A. Salomon, T. Barnighausen, and R. Atun, "Stopping Tuberculosis : a Biosocial Model for Sustainable Development," Lancet, vol. 386, no. 10010, pp. 2354–2362, 2015.

M. J. Saunders et al., "A Score to Predict and Stratify Risk of Tuberculosis in Adult Contacts of Tuberculosis Index Cases: a Prospective Derivation and External Validation Cohort Study," Lancet Infect. Dis., vol. 17, no. 11, pp. 1190–1199, 2017.

J. d. A. L. Batista et al., "Incidence and Risk Factors for Tuberculosis in People Living with HIV: Cohort from HIV Referral Health Centers in Recife, Brazil," PLoS One, vol. 8, no. 5, 2013, doi: 10.1371/journal.pone.0063916.

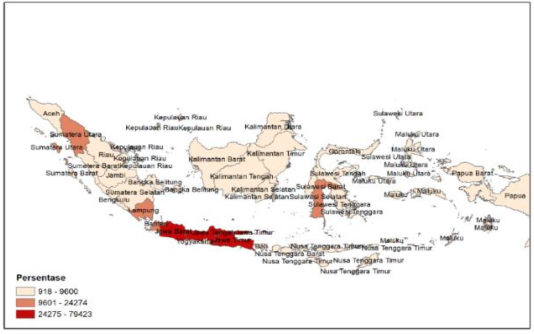

Figure. 1. Maping of Tuberculosis Cases in Indonesia

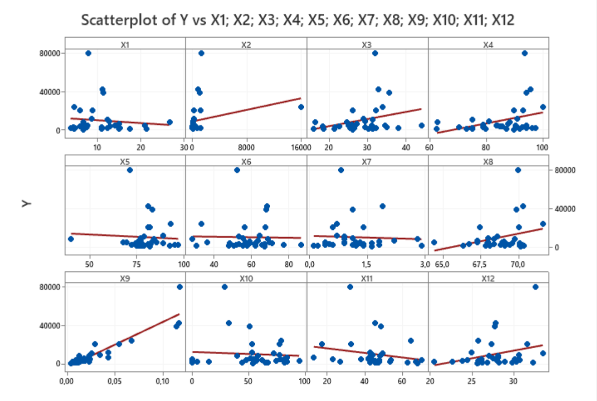

Figure. 2. Patterns of Relationship Between Variable

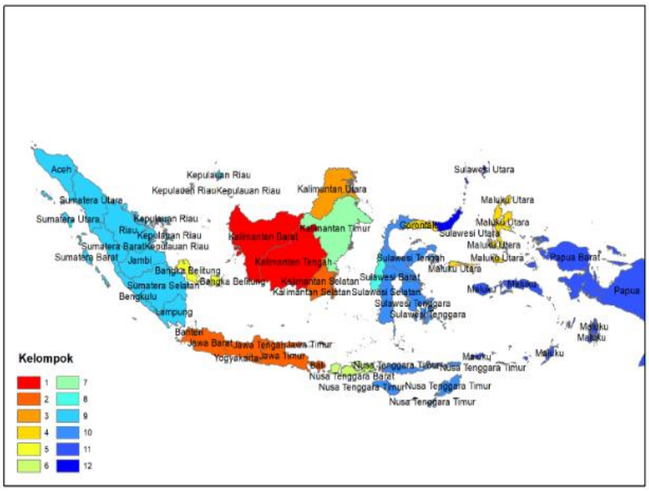

Figure. 3. *Results of Provincial Grouping with the*